

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 12:34:11 ; Search time 2954.29 Seconds
(without alignments)
13199.933 Million cell updates/sec

Title: US-09-357-273A-1

Perfect score: 3629
1 ccgagctcgcagcgcgcgtca.....caggagccagcccgaaatc 3629

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1019.4	25.1	1958	12 AK018505	Mus muscu
2	554.2	18.3	680	11 BG080244	H3051P08-
3	499.8	13.8	649	11 BE910769	601661819
4	454.8	12.5	823	11 BG083506	H3089H09-
5	447.4	12.3	482	11 BF846361	PM1-EN006
6	431.2	11.9	638	11 BF711744	PK0
7	333.4	9.2	380	11 BF846446	PM1-EN006
8	293.6	8.1	920	10 BE736116	601305874
9	277	7.6	279	11 BF868869	IT5-ET011
10	263.6	7.2	544	11 BG070677	H3089H09-
11	254	7.0	587	11 BF582789	602100922
12	241	6.6	556	10 AW642631	cm20g11.w

13	227.8	6.3	449	10 AA413881	AA413881
14	227.2	6.3	512	10 A1733735	A1733735
15	227.2	6.3	539	10 A1733738	A1733738
16	216.2	6.0	565	10 AW809457	MR4-ST012
17	212.4	5.9	677	11 BF964654	602267732
18	200	5.5	342	10 AA794000	AA794000
19	189.6	5.2	740	11 BF582794	602100927
20	189.2	5.0	476	10 AA088547	A2600154
21	180.6	4.9	539	11 B1394483	PM1-EN006
22	178.6	4.9	539	10 B1394483	PM1-EN006
23	151.2	4.2	421	10 A1734961	A1734961
24	151.2	4.2	422	10 A1734969	A1734969
25	139.4	3.8	348	10 AW846947	RC3-CT019
26	137.6	3.8	426	10 AU016754	AU016754
27	136.2	3.8	775	13 CANS02PEK	AL207893
28	133.4	3.6	170	11 BF817483	RC5-CT014
29	114.6	3.2	403	10 AA619271	AA619271
30	114.6	3.2	531	13 A2294936	RCPT-23-9
31	114.6	3.2	1101	13 CANS04U08	AL308249
32	114	3.1	561	11 BG710600	PM1-EN006
33	104.2	2.9	133	11 BF995998	MR1-GN017
34	104	2.9	780	11 BG784666	SEAUMC004
35	100.4	2.8	571	13 A2626875	1M0467D19
36	99	2.7	482	13 A2080607	RCPT-23-3
37	98	2.7	948	13 CANS0011B	AL074608
38	93.6	2.6	322	10 A1986903	rs19a12.Y
39	92.8	2.6	561	13 A0738301	HS-2260.B
40	92	2.5	999	13 CANS06MSD	AL405875
41	86	2.4	1134	13 CANS02345	T3 end of
42	85	2.3	348	13 B51730	AL179006
43	82.4	2.3	339	10 AA755189	B51730
44	82.2	2.3	283	10 AA102368	AA755189
45	81.2	2.2	721	13 A2711940	RCPT-24-8

ALIGNMENTS

RESULT 1	AK018505	1958 bp	mRNA	HTC	05-JUL-2001
LOCUS	AK018505				
DEFINITION	Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:9030414B18, full insert sequence.				
ACCESSION	AK018505.1 GI:12858236				
VERSION	1				
KEYWORDS	CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male colon cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.				
AUTHORS	1 (bases 1 to 1958)				
TITLE	Carninci, P. and Hayashizaki, Y.				
JOURNAL	High-efficiency full-length cDNA cloning				
MEDLINE	Methods in enzymology. 303, 19-44 (1999)				
PUBMED	99279253				
REFERENCE	2 (bases 1 to 1958)				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to				
JOURNAL	prepare full-length cDNA libraries for rapid discovery of new genes				
MEDLINE	Genome research. 10 (10), 1617-1630 (2000)				
PUBMED	20499374				
REFERENCE	3 (bases 1 to 1958)				
AUTHORS	Kono, H., Akiyama, J., Nishikawa, S., Nakamura, S., Sasaki, N., Carninci, P., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiiwa, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,				

Accession	Sequence	Length
Db 1091	TCGAGCAGACACCTCTTCTTGGTGGAAAGCCCCGACAGACATGCGCTACCATTTGAGG	1150
Oy 1080	caagagggaatbfygtatataacggcccgacggagcufcaagttatgaccggaactcoaaag	1139
Db 1151	CAAGGAGAGTGTGTATCTCTCCAGCACAGACTTCAAGTTTGACCTGGACTDAAAGG	1210
Oy 1140	caagaaacagctcaactactltgagaatctactgctcttcgatalagaaacacatlgaaaccc	1199
Db 1211	GAAAGCGAACCTGAACTACTTGGAGAAATCACTGGCTTCATCAGAGACACCATGAAACTCC	1270
Oy 1200	actgtcgcgtctaaccaagatgctltgagaagattcccaacaatcaccaccaaatcggga	1259
Db 1271	TCTGTCTGCATCCACCAAGATGCTGGAGAGATTTCCTAACACACTCGCCCAAAACATCGGA	1330
Oy 1260	aaatgcatctccgtcgtgattccagaagaaaagagcctltgagaagtta	1306
Db 1331	AAATGTATTTCTCTCTGATTCAGAAAAAGAGACTTTGAGGAAGGGA	1377

RESULT	2
BG080244	
LOCUS	BG080244 680 bp mRNA EST 26-JAN-2001
DEFINITION	H305ID08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cdna clone.
ACCESSION	H305ID08 5' mRNA sequence.
VERSION	BG080244
KEYWORDS	BG080244.1 GI:12562812
SOURCE	EST.
ORGANISM	house mouse. Mus musculus

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 680)	Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jarađat,S.A., Tanakawa,T.S., Carter,M.G. and Ko,M.S.H.	Verification and initial annotation of NIA mouse 15K cDNA clone set	Unpublished (2001)	Other_ESTs: H3051D08-3

National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@ngsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit <http://1gsun.grc.nia.nih.gov/cDNA/15x.html> for details.
Plate: H3051 row: D column: 08
Seq primer: -21M13 Reverse
High quality sequence stop: 680
POLYA-No.

```

FEATURES
source
location/Qualifiers
1. .680
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="H3051D08"
/clone_1ib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stages="Clones arrayed from a variety of cDNA
libraries"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI. This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5 embryos
, and E12.5 female mesonephros/gonad) and one newborn
ovary cDNA library. Average insert size 1.5 kb. All
source libraries are cloned unidirectionally with Oligo(dT
)-Not primers. References include: (1) Genome-wide
expression profiling of mid-gestation placenta and embryo/ov
using a 15,000 mouse developmental cDNA microarray,
Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
Large-scale cDNA analysis reveals phased gene expression

```

patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978.

Query Match	15.3%	Score 554.2;	DB 11;	Length 688;
Best Local Similarity	89.5%;	Pred. No. 1.7e-106;		
Matches 607;	Conservative 0;	Mismatches 69;	Indels 2;	Gaps 1;

[illegible]

Db	146	GAAGACTGTATGAGGAACCCCTACACGGGTGACATCTTTTCTGTGACAGGTGGCTCTTT	205
Qy	2389	tactacgtgtgtcctcttgaggcagccacccttttggcaatgctcctgcagcgagcccaac	2448
Db	206	TACTACGTAAATCTCTTGAGGGCAGCCACCCCTTTTGGCAAGTCCCTGCAGGGCAGGCCAAC	265
Qy	2449	atcttccttggtgtgcctcgaagccttgactgctgtcttcaccacagagaagacaagaacgtcatl	2508
Db	246	ATCTCTCTGGGTGGCTGGACGCTTGACCTGTCTTGACACCCAGAGGACAGAAAGACGTATT	325
Qy	2509	gcacagagaa-ctgtatagagaagaatgattgtcgatgtatgataccctcaagaagccctcta	2562
Db	326	GCACGTGAACCTGTGATAGAGAAAGATGATTCGCATGTTCTTCAGAAACGCCCTTCA	380
RESULT	8		
LOCUS	BE736116		
DEFINITION	601305874F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3640266 5',		
	mRNA sequence.		
ACCESSION	BE736116		
VERSION	BE736116.1	GI:10150108	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 920)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .		
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D.		

```

FEATURES
Source
    147 a      308 c      286 g      179 t
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/ELN1 at:
    http://image.llnl.gov
    Plate: LICM34 row: c column: 19
    High quality sequence stop: 657.
    Location/Qualifiers
        1..920
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:3640266"
        /clone_lib="NTH MGC 39"
        /tissue_type="adenocarcinoma"
        //host="DH10B (phage-resistant)"
        /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
        Site_2: EcoRI; cDNA made by oligo-dT priming.
        Directionally cloned into EcoRI/XhoI sites using the
        following 5' adaptor: GGCACGAG(G). Library constructed
        by Ling Hong in the laboratory of Gerald M. Rubin
        (University of California, Berkeley) using ZAP-cDNA
        synthesis kit (Stratagene) and Superscript II RT (Life
        Technologies)."

```

Query Match	8.1%	Score 293.6	DB 10	Length 920
Best Local Similarity	65.3%	Pred. No. 1.4e-51		
Matches 462, Conservative	0	Mismatches 244	Indels 2	Gaps 2

[illegible]

OY	2217	caaatcccaacttcttgctctcgaagaagctggaagtgaggagacaacagtttcaagcgc	2276
Db	128	ggtgctctcAACACTTGGGCTCTGcCAAGAAAGCTGCTGCTGGCGCTGTAGCTTCAAGCT	187
OY	2277	ccgaatctgggttgctctgacacaaagcttgatctgcctcagaagatgtctgaagcgaactg	2336
Db	188	CCACTCCGGCATCCCGGGCAAGGAGGCTGGATGGGGCCGAGCTTCTGACACTCTCCGC	247
OY	2337	taaggaagaaccttaacctacaacgltgagacatcttctcgaagtcgltcttctactacgt	2396
Db	248	ACCAGACAGTCTTACCAAGCGCTGGAGCATCTTCTCTGAGGCTGGTGTCTACTAGCT	307
OY	2397	ggtctctgaagcgcaagccaccttlttgcaagttccctgcgaagcgcaagccaatactctc	2456
Db	308	GCCTTCTGTGGGCGCCACCCCTTTTGGAGACAGTCTTTTATGCGACAGCAAAATCATCTCAC	367
OY	2457	gggtgcctgcagccttctacgcgtctgcacccagagaagacgaagaagctatctcagcaga	2516
Db	368	AGGGGCTTCCGTGTGGCTCACCCTGGAGGAAGAGTCCAGCAGCAAGTGGTTGCCGGGA	427
OY	2517	atgtatagaagaagtgtatltgcgatgtgattcttcgaagaacgcccctcagcgaaacgactgt	2576
Db	428	CTGTGTTGAGACCATGTTTGAAGCCACATGCCGAGCCACGCCCTGTGCCCCCAGTGGCT	487
OY	2577	caaacacccgctctctctgtagagccttagaagaagcagctccagttcttccagagctgtagcga	2636
Db	488	GGCCCCACCCCTTCTTTTGGAGCAGAGCAGCAAGTCAACTCTTCTTCCAGAGAGTCAAGTGA	547
OY	2637	cagatagaagaagaatcccttgtagtgccgagatcgtgaagcagttagtagagagggcgag	2696
Db	548	CTGGCTGGAGAGGATGCCAGCGAGGACCCCTGTGTGAGGGCACTGGAGCGGAGGCTG	607
OY	2697	agccgtgtgtagaagtgtgactgcggaagaaatcaatctgacccctccacaagactctcg	2756
Db	608	CGCAGTGGTCCGGGACAACTGGCAGCAGACATCTCCATATGCCGTCC - GACAGATCTGAG	666
OY	2757	taaatcagaacctataaagtggttctgtcagaagatctctcccgagc	2804
Db	667	AAAAGTCCGGTCTTAT - TAGGGGCTCTAGTGCAGAGCTCCTCCGGCG	713

	RESULT	9		
	BF686869/c			
LOCUS	BF686869	279 bp	mRNA	EST 17-JAN-2001
DEFINITION	T15-EOTU19-181000-17S-cl1 ET0119 Homo sapiens CDNA,			mRNA sequence
ACCESSION	BE686869			
VERSION	BF686869.1	GI:12259103		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 279)			
AUTHORS	Dias Neto,E., da Silva Correea,R., Verjowski-Almeida,S., Britones,M.R. Nagal,M.A., da Silveira,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.P., Metsckner,A., Balci,G.S., Simpson,D.H Brustein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,I.R.R., Reis,L.F., de Souza,S.J. and Simpson A.J.			
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7),	3491-3496	(2000)
MEDLINE	20202663			
COMMENT	Contact: Simpson A.J.G.			

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel.: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=IL5-ct2-IL5-ET0119-181000-175-cl1ct3=2000-10-18ct4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 241.

FEATURES

Source

1. .279

1. .279

BASE COUNT	61 a	83 c	62 g	71 e	2 others
ORIGIN					

Query Match	7.68;	Score 277;	DB 11;	Length 279;
Best Local Similarity	99.3%;	Pred. NO. 3.2e-48;		
Matches 277; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	3216	gtgacccacagggaaactggggagaaagtctgccccaaagacctaagtcatagtctgc	3275
Db	279	GTGACCTACAGGGAACTGGGAATACTGCTGCNCNCANNAACCTTACGCTATATATGTCTGC	220
QY	3276	aaaggaggccctcagagacagcgcgagtagcaaccacacalactactgataaaacttgc	3335
Db	219	AAAGGAGGCGCTCAGAGACGCCGCGATGACACCCACCACTCTACATGATAAATCTTGC	160
QY	3336	ttcagactttttaaattccgcgtctaattgcactcagagcctttcagaaaggagaaaga	3395
Db	159	TTTCAGACTTTTAAATTCCTGCTTAATGTGACGTACAGAGCCTTTCAGGAAGGAGAGGA	100
QY	3396	gggaatcgtacatttctgcttgcgtctggagacagctagagcttagatgacccaagtcagc	3455
Db	99	GGGAATCGTACATTTTGCTTGCGCTGCGGACAGCTAGAGCTGAGATGACCAACAACTACAGC	40
QY	3456	cttcactcgtgagaccggaattcgaagaagtcgggagatgctctga	3494
Db	39	CTTTCACGTGAGACCGGAATTTGAGAGGTGGGGGTGCTGTA	1

RESULT 10
RG070677/c

ACCESSION BG070677
VERSION BG070677.1 GI:12553246
KEYWORDS EST.
SOURCE house mouse.

REFERENCE	TITLE
Authors	
1 (bases 1 to 544)	
Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanakke, T.S., Carter, M.G. and Ko, M.S.H.	
Verification and initial annotation of NIA mouse 15K cDNA clone set	
Unpublished (2001)	
Other_ESTs: H3089H09-5	
Comment	

Seq primer: -21M13 Forward
High quality sequence stop: 544
POLYA=Yes.

FEATURES

Source

1. .544

1. .544

BASE COUNT	126 a	117 c	139 g	162 f
------------	-------	-------	-------	-------

Query Match	7.28;	Score 262.6;	DB 11;	Length 544;
Best Local Similarity	86.88;	Pred. NO. 4.2e-45;		
Matches 289;	Conservative 0;	Mismatches 44;	Indels 0;	Caps 0;

QY	997	cacgagggggtgtcgtctgtgtgtcccgcgagacacacttcttctgtcgtgaaggcccg	105b6
Db	544	CATGAGGGGGTTCCTGTCTGTGCTCTCAAGGACAGACTCTTCTCTGGAAGGCCCCAG	485
QY	1057	actgaatggctaccacatcggtagcaaggaggaggtgtgtatacggcccgacgagctc	111b6
Db	484	ACAATATGGCTCACCATTTGAGAGACAAAGAGAGGTGTGTATATCTCCAGCACAACTTC	425
QY	1117	aagttgtatcccgagctcaaaagcaagaacaagctcaactactttaggaatctcgtct	117b6
Db	424	AAGTTTATACCTTGAGCTACACAAAGGAGAGACCAAGCTAACTACTTGAAGATTAACGCGTT	365
QY	1177	ctgataaggacacatgaanaccacactgtctgtcgtctacaaagctcgtgaagatctcc	123b6
Db	364	CTCATAGGACACACATGAAACTCTCTGTCTGCATCCACCAAGATGCTGGAGAGATTTCCT	305
QY	1237	aacacatctaccacaacatcogtgggaatatgtatctctcgtcgtactcggagaaaaagcctt	129b6
Db	304	AACAACCTGGCCCAAACTCAGAGAAATGTGATTCTCTGTGATTCTAGAAAAGAGCGTTT	245
QY	1297	gaggaagtatacaacctgtgtctgacagacttca	1329
Db	244	GAGGAAAGATCAAACTCAGATTGTCTCAGACTTGA	212

RESULT	11								
LOCUS	B5582789								
DEFINITION	B5582789	587 bp	mRNA		EST	12-DEC-2000			
ACCESSION	602100922E1	NCL_CGAB_CO24	MUS	MUSCULUS	CDNA	clone	IMAGE:4224148	5'	
VERSION	B5582789.1	GI:11656507							

